DOCUMENT RESUME

ED 388 889 CG 025 958

AUTHOR Stiggins, Richard J.
TITLE Sound Performance Assessments in the Guidance Context. ERIC Digest.
INSTITUTION ERIC Clearinghouse on Counseling and Student Services, Greensboro, NC.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO EDO-CG-95-9
PUB DATE 95
CONTRACT RR93002004
NOTE 4p.
PUB TYPE Information Analyses - ERIC Clearinghouse Products (071)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Educational Assessment; Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; Evaluation Problems; Evaluation Research; Evaluation Utilization; Nongraded Student Evaluation; Performance Tests; School Counselors; School Guidance; *Student Evaluation; Teacher Expectations of Students; *Testing
IDENTIFIERS ERIC Digests

ABSTRACT

Not since the development of the objective paper and pencil test early in the century has an assessment method hit the American educational scene with such force as has performance assessment methodology in the 1990s. Performance assessment relies on teacher observation and professional judgment to draw inferences about student achievement. Educators have begun to embrace the reality that some targets, like complex reasoning, skill demonstration and product development require the use of subjective, judgmental means of assessment. This digest provides a summary of attributes of sound assessments and the rules of evidence for using them well. Various ways to take advantage of this information are presented. The following aspects are detailed: The Basic Methodology; Sound Performance Criteria; Sound Performance Exercises; Effective Scoring and Recording; and Performance Assessment in the Guidance Context. (JBJ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 388 889

ERIC DIGEST EDO-CG-95-9

Sound Performance Assessments in the Guidance Context

by Richard J. Stiggins

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

CG025958

ERIC CLEARINGHOUSE ON Counseling And Student Services

ERIC DIGEST

1995 EDO-CG-95-9

Sound Performance Assessments in the Guidance Context

Richard J. Stiggins

Not since the development of the objective paper and pencil test early in the century has an assessment method hit the American educational scene with such force as has performance assessment methodology in the 1990s. Performance assessment relies on teacher observation and professional judgement to draw inferences about student achievement. The reasons for the intense interest in an assessment methodology can be summarized as follows:

During the 1980s important new curriculum research and development efforts at school district, state, national and university levels began to provide new insights into the complexity of some of our most valued achievement targets. We came to understand the multidimensionality of what it means to be a proficient reader, writer, and math or science problem solver, for example. With these and other enhanced visions of the complex nature of the meaning of academic success came a sense of the insufficiency of the traditional multiple choice test. Educators began to embrace the reality that some targets, like complex reasoning, skill demonstration and product development, *require*—don't merely permit—the use of subjective, judgmental means of assessment. One simply cannot assess the ability to write well, communicate effectively in a second language, work cooperatively on a team, and complete science laboratory work in a quality manner using the traditional selected response modes of assessment.

As a result, we have witnessed a virtual stampede of teachers, administrators and educational policy makers to embrace performance assessment. In short, educators have become as obsessed with performance assessment in the 1990s as we were with the multiple choice tests for 60 years. Warnings from the assessment community (Dunbar, Kortez, and Hoover, 1991) about the potential dangers of invalidity and unreliability of carelessly developed subjective assessments not only have often gone unheeded, but by and large they have gone unheard.

Now that we are a decade into the performance assessment movement, however, some of those quality control lessons have begun to take hold. Assessment specialists have begun to articulate in terms that practitioners can understand the rules of evidence for the development and use of high quality performance assessments (e.g. Messick, 1994). As a result, we are well into a national program of research and development that builds upon an ever clearer vision of the critical elements of sound assessments to produce ever better assessments (Wiggins, 1993).

The purpose of this digest is to provide a summary of those attributes of sound assessments and the rules of evidence for using them well. The various ways the reader might take advantage of this information also are detailed.

The Basic Methodology

The basic ingredients of a performance assessmentmay be described in three parts (Stiggins, 1984): (1) the specification of a performance to be evaluated, (2) the development of exercises or tasks used to elicit that performance and (3) the design of a scoring and recording scheme for results. Each contains subelements within it.

For example, in defining the performance to be evaluated, assessment developers must decide where or how evidence of academic proficiency will manifest itself. Is the examinee to demonstrate the ability to reason effectively, carry out other skills proficiently or create a tangible product? Next, the developer must analyze skills or products to identify performance criteria upon which to judge achievement. This requires the identification of the critical elements of performance that come together to make it sound or effective. In addition, performance assessors must define each criterion and articulate the range of achievement that any particular examinee's work might reflect, from outstanding to very poor performance. And finally, users can contribute immensely to student academic development by finding examples of student achievement that illustrate those different levels of proficiency.

Once performance is defined, strategies must be devised for sampling student work so skills or products can be observed and evaluated. Examinees might be presented with structured exercises to which they must respond. Or the examiner might unobtrusively or opportunistically watch performers during naturally occurring classroom work in order to derive evidence of proficiency. When structured exercises are used to elicit performance,they must spell out a clear and complete set of performance responsibilities for examinees. In addition, the examiner must include in the assessment enough exercises to sample the array of performance possibilities in a representative manner that is large enough to lead to confident generalizations about examinee proficiency.

And finally, once the desired performance is described and exercises have been devised, procedures must be spelled out for making and recording judgments. These scoring schemes, somestimes called rubrics, help the evaluator translate judgments of proficiency into ratings. The assessment developer must select the level of detail to be reflected in records, the method of recording results, and who will be the observer and rater of performance.

Sound Performance Criteria.

Quellmalz (1993) offers a set of specific guidelines for the development of quality performance criteria. These reflect important aspects of skill demonstration that judges are to look for and evaluate—they represent important attributes of quality products.They are devised through a thoughtful analysis of samples of high quality performance and comparison to samples of inferior performance. Out of this comparison come an un-

derstanding of the keys to academic success in the context for which the assessment is designed. Quellmalz advises us that criteria should: be significant, specifying important performance components; represent standards that would apply naturally to determine the quality of performance when it typically occurs; be generalizable—that is, applicable to a class or tasks— not apply to only one task appropriate continuum from low-to high-level achievement; communicate clearly to and be able to understood by all involved in the performance assessment process, including teachers, students, parents and community; hold the promise of communicating information about performance quality that provides a basis for the improvement of that performance. (p, 320)

The attributes of quality performance that form the basis of judgement criteria should be couched in the best current thinking about the keys to academic success as defined in the professional literature of the discipline in question.

Sound Performance Exercises

Baron (1993) provides guidance in the development of sound exercises. These spell out the achievement to be demonstrated by the examinee, the conditions under which the demonstrations will take place and the criteria that will serve as the basis for evaluation of performance. In short, they focus the examinee sharply on the task at hand. Baron advises that these questions be used to determine exercise quality: when students prepare for my assessment tasks and I structure my curriculum and pedagogy to enable them to be successful on these tasks, so I feel assured that they will be making progress toward becoming genuine or authentic readers, mathematicians, writers, historians, problem solvers, etc.; do my tasks clearly communicate my standards and expectations to my students; are some of my tasks rich and integrative, requiring students to make connections and forge relationships among various aspects of the curriculum; do some of my tasks require that my students sustain their efforts over a period of time (perhaps even an entire term!) to succeed; do my tasks require self-assessment and reflection on the part of students; are my tasks likely to have personal meaning and value to my students; and do some of my tasks provide problems that are situated in real-world contexts and are they appropriate for the age group solving them?

Effective Scoring and Recording

The basis of the effective application of performance assessment methodology is thoroughly trained raters relying on sound performance criteria to observe and evaluate student responses to quality exercises (Stiggins, 1994). It is rarely the case that raters can automatically judge student performance merely as a matter of their prior professional development. Training—or at least a systematic verification of qualifications to rate performance—is essential in all contexts in which quality assessment results are the goal.

One test of the quality of ratings is interrater agreement. A high level of degree of agreement is indicative of objectivity of ratings. Another test of quality is consistency in a particular rater's judgments over time. Ratings should not drift but rather should remain anchored to carefully defined points on the scoring scale. A third index of performance rating quality is consistency in ratings across exercises intended to be reflective of the same performance—an index of internal consistency. When these standards are met, it becomes possible to take advantage of the immense power of this kind of assessment to muster concrete evidence of improvement in student performance over time.

There are three design decisions to be made by the performance assessment developer with respect to scoring schemes: the level of specificity of scoring, the selection of the record keeping method, and the identification of the rater. Scores can be holistic or analytical, considering criteria together as a whole or separately. The choice is a function of the assessment purpose. Purposes like diagnosing weaknesses in student performance that require a high resolution microscope require analytical scoring.

Recording system alternatives include checklists of attributes present or absent in performance, rating scales reflecting a range in performance quality, anecdotal records that describe performance or mental record keeping. Each offers advantages and disadvantages depending on the specific assessment context.

Raters of performance can include the teacher, another expert, students as evaluators of each other's performance or students as evaluators of their own performance. Again, the rater of choice is a function of context. However, it has become clear that performance assessment represents a powerful teaching toolwhen students play roles in devising criteria, learning to apply those criteria, devising exercises, and using assessment results to plan for the improvement of their own performance—all under the leadership of their teacher.

Performance Assessment in the Guidance Context

The ongoing guidance and counseling function in the school could bring student service personnel into contact with performance assessment methodology in three important ways. Very often, other education professionals regard counselors as sources of expertise in assessment and may bring request for opinions about the value of this methodology, or they may ask for help in the design and development of performance assessments.

Or counselors might be invited to serve as raters of student performance in specific academic disciplines. If and when such opportunities arise, through training is essential for all who are to serve in this capacity. If the teachers issuing this invitation have developed or gleaned from their professional literature refined visions of the meaning of academic success, have transformed them into quality criteria and provide quality training for all who are to observe and evaluate student performance, this can be a very rewarding professional experience. If these standards are not met, it is wise to urge (and perhaps help with) a redevelopment of the assessment.

The third and final contact for counselors is as an evaluator of students within the context of the guidance function, observing and judging academic or affective student characteristics. In this case, the counselor will be both the developer and user of the assessment and must know how to adhere to the above mentioned standards of assessment quality.

For all of these reasons, it is advisable for school guidance and counseling personnel to understand when this methodology is likely to be useful and when it is not and how to design and develop sound performance assessments.

References

Baron, J.B. (1991). Strategies for the development of effective performance assessment exercises. Applied Measurement in Education, 4(4), 305-318.

Dunbar, S.B., Kortez, D.M., & Hoover, H.D. (1991). Quality control in the development and use of performance assessments. Applied Measurement in Education, 4(4), 289-304.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. Educational Researcher, 23(2), 13-23.

Quellmalz, E.S. (1991). Developing criteria for performance assessments: The missing link. Applied Measurement in Education, 4(4), 319-332.

Stiggins, R.J. (1994). Student-Centered Classroom Assessment. Columbus, OH: Macmillan.

Wiggins, G.P. (1993). Assessing Student Performance. San Francisco, CA: Jossey-Bass.

Rick Stiggins is Director of the Assessment Training Institute in Portland, Oregon.

ERIC Digests are in the public domain and may be freely reproduced and disseminated. This publication was funded by the U.S. Department of Education, Office of Educational Research and Improvement, Contract No. RR93002004. Opinions expressed in this report do not necessarily reflect the positions of the U.S. Department of Education, OERI, or ERIC/CASS.

For information on other ERIC/CASS products and services, please call toll-free (800) 414-9769 or (910) 334-4114 or fax (910) 334-4116 or write ERIC/CASS, School of Education, University of North Carolina at Greensboro, Greensboro, NC 27412.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

NOTICE

REPRODUCTION BASIS

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☒ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").